

Turnover Rate of Popularity Charts in Neutral Models

T.S. Evans^{1,2} and A. Giometto¹

¹ *Institute for Mathematical Sciences, Imperial College London, London, SW7 2PG, UK*

² *Theoretical Physics, Imperial College London, SW7 2AZ, UK*

It has been shown recently that in many different cultural phenomena the turnover rate on the most popular artefacts in a population exhibit some regularities. A very simple expression for this turnover rate has been proposed by Bentley et al. [10] and its validity in two simple models for copying and innovation is investigated in this paper. It is found that Bentley's formula is an approximation of the real behaviour of the turnover rate in the Wright-Fisher model, while it is not valid in the Moran model.

I. INTRODUCTION

In genetics, neutral models in which the different variations provide no intrinsic advantage play a central role. Classic examples are the Wright-Fisher and Moran models [1, 2] which are Markov processes. It is not surprising that they are also particular limits of other statistical physics models such as Urn models and zero range processes [3, 4] or network rewiring models [12, 14] (see [12] for further examples of these models). One of the most interesting applications is to cultural transmission [5–9]. In this case the popularity of ‘artefacts’ with no intrinsic value (uniform fitness) varies because individuals change their choice of artefact by either copying the artefact chosen by another individual (inheritance) or by innovating by choosing a new artefact (mutation). Under this hypothesis a certain cultural trait becomes more popular than others simply through imitation and not because of an intrinsic benefit it provides. Despite its simplicity, these models can reproduce some of the features of real data sets, such as Neolithic pottery [5, 9], popular music charts [6], baby names [7, 9], patents [9] and dog breeds [8].

Much is known about these neutral models, including many exact results [1, 2]. However the context of cultural transmission throws up new and unanswered questions since they are of little practical use in other applications. In this paper we study popularity charts, the list of the y most popular artefacts at any one time, and ask how many artefacts enter or leave this list each time the chart is updated, the turnover rate z . This is motivated by the work of Bentley et al. [10] who find that in the Wright-Fisher model the turnover rate z of the top y chart is $z = \sqrt{\mu} \cdot y$ where μ is the innovation rate. This result is interesting because the turnover rate z is independent of the population size N and the square root dependence on μ is reminiscent of a random walk process that might be addressed with a theoretical analysis. It is of practical use as sometimes we only have access to data on the most popular artefacts and we may not have a useful sample of the whole population. In such situations it can be used to provide estimates of the model parameters from a data set [11]. The aim of this paper is to perform a comprehensive study of the turnover rate in the Wright-Fisher and Moran models.

Individual	t	t+1
1	A	$\xrightarrow{\text{copy } 3}$ B
2	A	$\xrightarrow{\text{copy } 2}$ A
3	B	$\xrightarrow{\text{copy } 1}$ A
4	A	$\xrightarrow{\text{innovate}}$ D
5	C	$\xrightarrow{\text{copy } 6}$ A
6	A	$\xrightarrow{\text{copy } 6}$ A

Top 3 chart:

Position	t	t+1
1 st	A	A
2 nd	B	$\xrightarrow{z=2}$ B
3 rd	C	D

FIG. 1. A simple representation of the Wright-Fisher model. The artefacts are labelled by letters. In this example two successive time steps are shown for a population of six individuals and the top three chart with a turnover of two ($N = 6$, $y = 3$, $z = 2$).

II. THE WRIGHT-FISHER MODEL

The model investigated here and in [10] is one of the family of Wright-Fisher models and is illustrated in Fig. II. In it each of N individuals is characterised by an artefact of no intrinsic value (e.g. a brand of shoes, a dog breed, a name, etc.). At each time step all individuals in the population are simultaneously assigned a new artefact. With probability $(1 - \mu)$ an individual will copy the artefact choice from the previous time step of an individual selected uniformly at random. Otherwise with probability μ an individual innovates by choosing a new artefact.

The analytical solutions of this model [2] show that the frequency of artefacts in a population is typically a power law with a cutoff (at least for $N \gg \mu N \gtrsim 1$) and this has been fitted to data on the frequency of various modern cultural variants [7–9].

Definition of turnover

Our definition of the turnover z in the top y chart, the list of the y most popular artefacts, is defined as the sum of the number of artefacts exiting the top chart plus the numbers of new artefacts entering the top chart at the same time step. This definition of turnover is slightly different from [10] where it is defined as the number of new artefacts that enter the top y chart relative to the previous time step. In most situations the difference between the two definitions is given by a factor of 2. Our definition is more informative for situations where a artefact exits the top chart by becoming extinct with no new artefacts entering it: in this configuration we have a turnover $z = 1$, while the definition in [10] would have $z = 0$. In our notation the result of [10] is that

$$z = 2 \cdot \sqrt{\mu} \cdot y \quad (1)$$

Bentley et al. [10] find this through numerical analysis of this Wright-Fisher model and also find support for this form in data for baby names and dog breeds.

Simulations of the model

During one simulation the model starts with every individual assigned to a unique artefact and is then first updated τ times. After reaching a steady state the frequency of every artefact in the population is computed at each time step and the top y chart is built using the *quicksort* algorithm for the next T steps. The temporal average \bar{z} of the turnover rate z is then computed by comparing two successive top y charts and recorded. To perform an ensemble average the model is rerun E times¹ and the ensemble average $\langle \bar{z} \rangle$ of \bar{z} is computed. This is the estimate of the turnover rate z that is stored for further analysis along with an estimate of the standard deviation in this measurement.

We started our simulations from a configuration where all the individuals had a different artefact. We checked that our simulations had reached a steady state by studying

$$F_2(t) = \frac{\langle k(k-1) \rangle}{N(N-1)}(t) = \frac{\sum_{k=0}^N k(k-1) \langle n(k,t) \rangle}{N(N-1)} \quad (2)$$

where $n(k,t)$ is the number of artefacts chosen by k individuals at time t and the symbols $\langle \dots \rangle$ indicate an ensemble average. This quantity can be shown analytically to evolve in time as

$$F_2(t) = F_2(\infty) + [F_2(0) - F_2(\infty)] \cdot (\lambda_2)^t \quad (3)$$

¹ E is chosen so that the error on z is not larger than 10%. To reduce computational times the model is run for τ time steps to reach a steady state only once. Successive iterations start from the last configuration of the previous ones. This procedure doesn't affect the results because the system is in a steady state.

with $\lambda_2 = (1 - \mu)^2(N - 1)/N$ [2] so that $\tau^{-1} \sim \ln(\lambda_2) \lesssim \min((2\mu)^{-1}, N)$ (a similar results hold for all eigenvalues). Given the values used in our simulations we chose $\tau = 4\mu^{-1}$ and ran for $T = 50 + \mu^{-1}$ time steps (50 time steps were added to ensure a minimum amount of time steps even for big values of μ).

Analysis

Motivated by [10] we fitted our data for z as a function of μ , y and N to the following form

$$z = d \cdot \mu^a y^b N^c \quad (4)$$

We looked at around 6000 different parameter values taken from the ranges $\mu \in [5 \cdot 10^{-5}, 0.115]$, $y \in [2, 1411]$ and $N \in [180, 3993]$ with the constraint that $y < N$. This largely extends the range of values studied in [10] which come from $\mu \in [2 \cdot 10^{-4}, 0.02]$, $y \in [5, 50]$ and $N \in [500, 4000]$. Estimates for coefficients a , b , c and d in equation (4) were obtained using a linear fit to the data for $\ln(z)$.

It is found that the turnover rate z exhibits two different behaviours in the two regions: $N\mu < 0.15 \cdot y$ and $N\mu > 0.15 \cdot y$, as can be seen in figure 2, with the transition between these two behaviours occurring around $N\mu \simeq 0.15 \cdot y$.

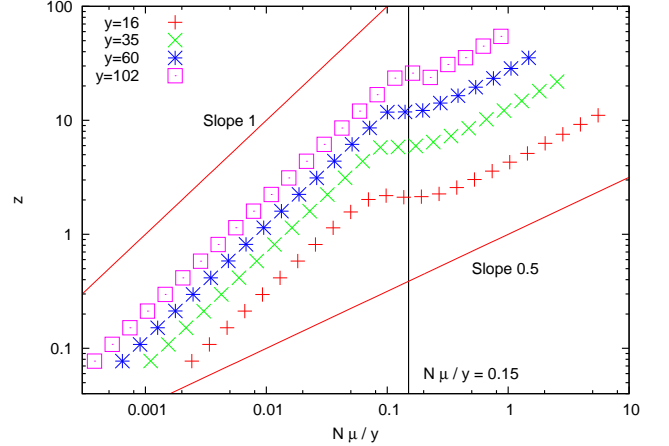


FIG. 2. Existence of a critical point $N\mu \simeq 0.15 \cdot y$. Plotted curves are for fixed values of y and N . Error bars are smaller than symbols.

Region $N\mu < 0.15 \cdot y$:

The observed behaviour of the turnover rate z in this region is: $z \propto \mu$ (see figure 2 and figure 3). Fitting the data points in this region with the functional form (4) the following set of values for the fitting parameters is

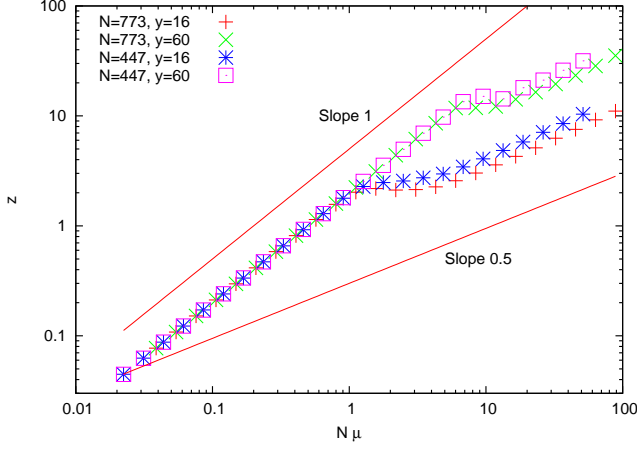


FIG. 3. Existence of a critical point $N\mu \simeq 0.15 \cdot y$. Plotted curves are for fixed values of y and N . In the region $N\mu < 0.15 \cdot y$ all curves collapse in $z = 2N\mu$. Error bars are smaller than symbols.

obtained:

$$\begin{aligned} a &= 0.99999(4), \quad b = -0.00004(7), \\ c &= 1.0003(2), \quad d = 1.997(2), \end{aligned} \quad (5)$$

that is to say:

$$z = 2 \cdot N\mu \quad (6)$$

within two standard deviations. Note that in this region z is independent on the top chart size y . This is best seen in figure 3, where the turnover rate is plotted against the product $N\mu$.

This behaviour of the turnover rate z can be explained in the following way: for $N\mu \ll y$ the average number of new artefacts that enter the population in one time step is lower than the top chart size y . In this configuration a very small number of artefacts survives in the population in the steady state. It is then likely that the total number of artefacts at a certain time t is lower than the top chart size y , as it is confirmed by observations (see figure 4). In one time step, then, the $N\mu$ new artefacts (on average) introduced in the previous time step (that are in the top y chart) are extinguished through copying, while on average $N\mu$ new artefacts enter the population and the top chart through innovation: the turnover rate z is then equal to $2 \cdot N\mu$. This mechanism is illustrated in figure 5.

Region $N\mu > 0.15 \cdot y$:

This is the region studied in [10] and we also find that the dependence of the turnover rate z on the innovation rate μ is very roughly $z \propto \mu^{\frac{1}{2}}$, as can be seen in figure 2. However we have also fitted the data in this region to the same functional form $z = d \cdot \mu^a y^b N^c$ (4), using a

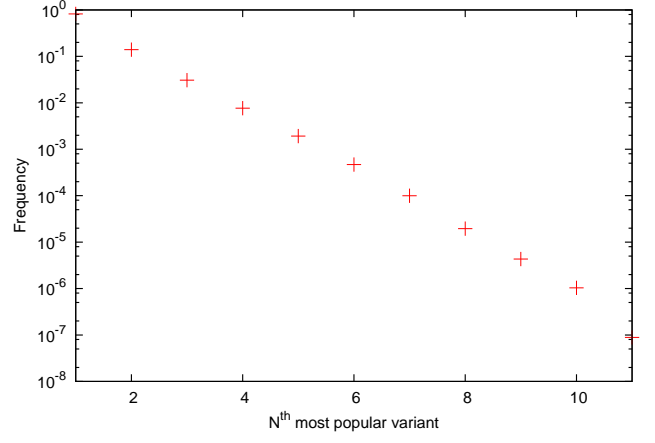


FIG. 4. Average frequency of artefacts (from the most popular one to the least popular). In the steady state there are no more than 11 artefacts in the population. With $N = 180$, $\mu = 0.001$ and $y = 20$ we have $N\mu \ll 0.15 \cdot y$ and the number of artefacts in the population (11) is smaller than the top chart size y (20).

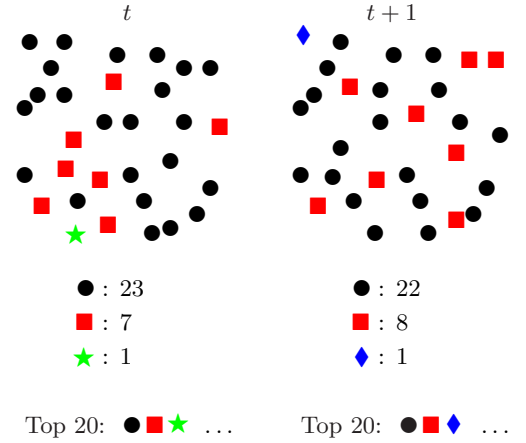


FIG. 5. A population of different symbols, the shape being the artefact. With $N\mu \ll y$ only a few artefacts survive in the population and the top chart has empty spots. New artefacts introduced through innovation (on average $N\mu = 1$ per generation) enter the top y chart but are extinguished in one time step, producing a turnover equal to $2 \cdot N\mu$.

linear fit to the logarithm of our parameters and z . The resulting values for the fitting parameters a , b , c and d are:

$$\begin{aligned} a &= 0.550(2), \quad b = 0.860(1), \\ c &= 0.130(2), \quad d = 1.38(2). \end{aligned} \quad (7)$$

These values are not statistically compatible with the proposed form in equation (1) for which $a = 1/2$, $b = 1$, $c = 0$ and $d = 2$. The dependence on μ and y is not so far off that proposed in [10] (a and b are 10% and -14% off the values in (1)) so the practical difference in

studying a real data set may be minimal. However we find a significant dependence on N , even with this small power of c we have a 50% variation in z over the range $N \in [180, 3993]$. This dependence on the population size N is clearly seen in figure 6.

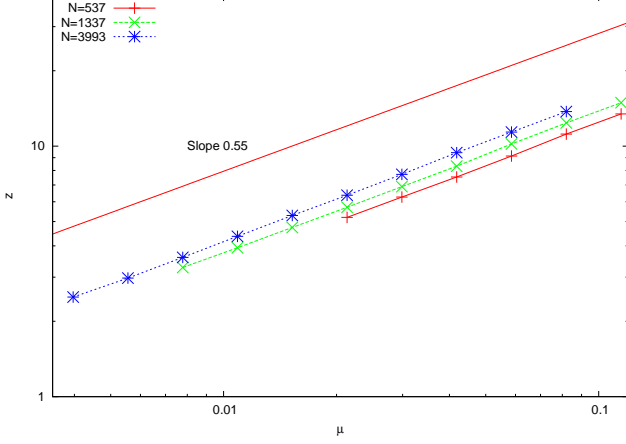


FIG. 6. Turnover rate z vs μ for the Wright-Fisher model in the region $N\mu > 0.15 \cdot y$. There is a clear dependence of z on the population size N . Error bars are smaller than symbols.

As a final illustration we plot the turnover rate z against the form (4) using our best fit values (7) to create a data collapse. As it can be seen in the figure, collected data lay on the diagonal $z = \mu^a y^b N^c$. The same plot using the form suggested in [10] is presented in figure 8.

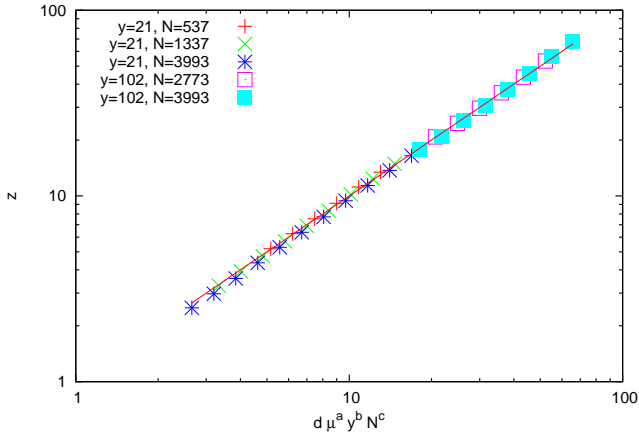


FIG. 7. Turnover rate z vs $d \cdot \mu^a y^b N^c$ for the Wright-Fisher model in the region $N\mu > 0.15 \cdot y$. Values used were $a = 0.550$, $b = 0.860$, $c = 0.130$ and $d = 1.38$, the best fit values found (7). Error bars are smaller than symbols.

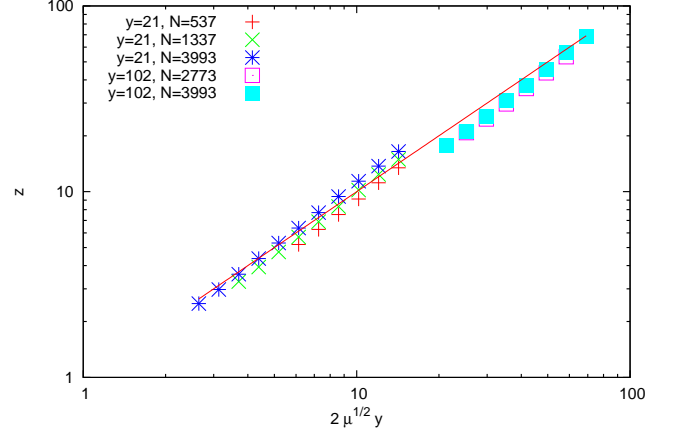


FIG. 8. Turnover rate z vs $2\sqrt{\mu}y$ for the Wright-Fisher model in the region $N\mu > 0.15 \cdot y$. Error bars are smaller than symbols.

Residuals

Residuals from the fit are shown in figure 9. While the peaks for low value of the product $N\mu$ are related to data near the critical point $N\mu \simeq 0.15 \cdot y$, there is an evident systematic deviation from the proposed functional form (4) for $N\mu \gg 0.15 \cdot y$.

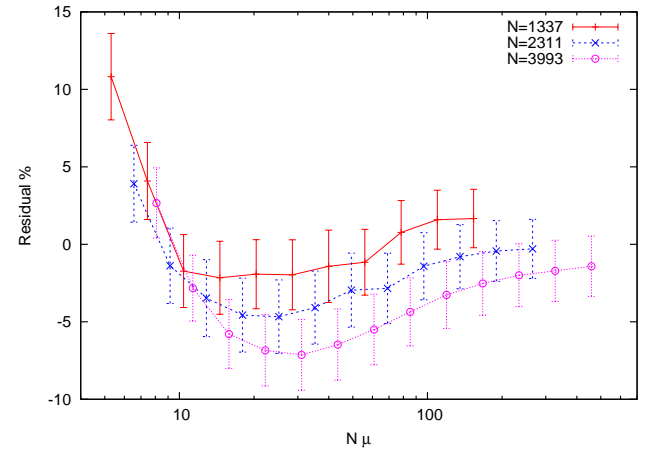


FIG. 9. Residuals from equation (4) vs $N\mu$. There is a systematic deviation from the proposed functional form. The deviations for small $N\mu$ are related to data near the critical region $N\mu \simeq 0.15 \cdot y$.

The minimum point of the residuals curves lies in $N\mu \simeq y$ for all values of N and y and increases in absolute value for increasing population size N , causing a systematic drift from equation (4).

We tried many different functional forms and the best

fits we found were obtained with the following:

$$\text{Res}(\mu, y, N) = R_0 \cdot \left[1 - e^{-f \cdot (\mu N - y)} \right]^2 + R_{\min}$$

$$R_{\min}(y, N) = A \cdot y + B \cdot N$$

where R_{\min} is the value of the residual at the minimum. The best fit values for the parameters are $R_0 = 0.062(4)$, $f = 0.0096(4)$, $A = 0.00034(5)$ and $B = -0.0000193(8)$. Combining this expression for the residuals with the original form (4) the following function might be used to describe the behaviour of the turnover rate z of the Wright-Fisher model in the region $N\mu > 0.15 \cdot y$:

$$z = d \cdot \mu^a y^b N^c \cdot [1 + \text{Res}(\mu, y, N)] \quad (8)$$

Residuals from equation (8) are plotted in figure 10. The

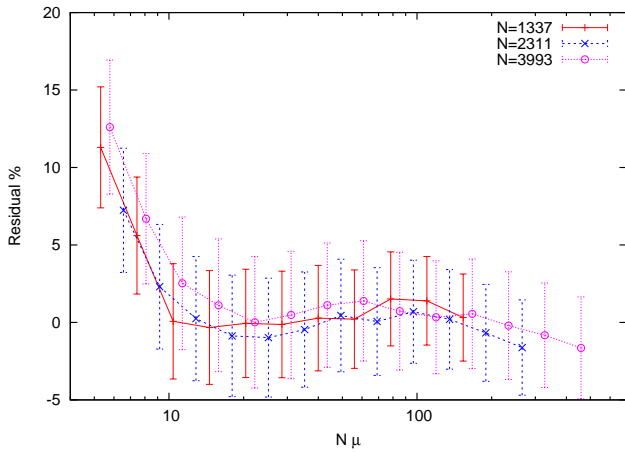


FIG. 10. Residuals from equation (8) vs $N\mu$. No systematic deviation from the proposed functional form (8) can be identified. The deviations for small $N\mu$ are related to data near the critical region $N\mu \simeq 0.15 \cdot y$.

functional form (8) can now reproduce data within 2% for $N\mu \gg 0.15 \cdot y$ and there is no systematic drift in $N\mu \simeq y$ for increasing values of N .

However this form (8) is not very satisfactory because of the high number of parameters used to fit the data and because the range of applicability of equation (8) is strongly limited, due to the negative value of B .

Large populations

The numerical studies of the Wright-Fisher model in [10] have only a small number of individuals $N \leq 4000$, whereas the real data sets are often much larger e.g. around 10^6 births per year in [7], 6×10^5 dog breed registrations per year in [8], the number of albums released in the US (data used in [6, 10]) is hard to estimate but the Recording Industry Association of America have suggested that major labels release around 7,000 new CDs every year.

Unfortunately all data sets on which this model has been applied considered population of millions of individuals and this would take too long to simulate. The biggest population size that we have been able to simulate is of the order of 100000 individuals. Results are listed in table I.

μ	y	N	z	$z/(2y\sqrt{\mu})$	z/z_{small}	z/z_{large}
0.0012	200	100000	11.9(1)	1.16	1.22	1.06
0.0024	200	100000	17.6(1)	1.11	1.21	1.05
0.0012	400	100000	24.3(1)	1.14	1.09	0.95
0.0024	400	100000	33.3(2)	1.18	1.16	1.03
0.0012	200	120000	12.1(1)	1.15	1.23	1.06
0.0024	200	120000	18.0(1)	1.09	1.21	1.05
0.0012	400	120000	23.8(1)	1.16	1.13	0.99
0.0024	400	120000	33.6(2)	1.17	1.18	1.03
0.0012	200	144000	12.2(1)	1.14	1.25	1.07
0.0024	200	144000	18.3(1)	1.07	1.22	1.05
0.0012	400	144000	23.6(1)	1.17	1.17	1.02
0.0024	400	144000	34.0(2)	1.15	1.19	1.04

TABLE I. Simulation results for big population size N . The last three columns give the ratios of the measured value of the turnover z divided by one of the fitted forms $z = dy^a \mu^b N^c$, respectively $z = 2y\sqrt{\mu}$ of [10] and (1), z_{small} best fit for small N of (7), and finally z_{large} best fit for large N of (9). All results and best fits are for the $\mu N \gtrsim 0.15y$ region.

A surprising result is that for the data in table I the simple form suggested in [10], $z = 2\sqrt{\mu y}$, is a better estimate for the turnover rate (typically a 15% overestimate) than our best fit to $z = dy^a \mu^b N^c$ of (4) using the data for small N but large μN (typically around 20% overestimate). A factor that might produce this unexpected result is that in all real situations where the model has been tested and in the data in table I the size of the top chart is much smaller than the population size: $y \ll N$. We have therefore performed a new data collection in this region (and above the critical point $N\mu \simeq 0.15 \cdot y$).

Approximately 1200 data points have been collected in the range $\mu \in [0.001, 0.13]$, $y \in [5, 56]$, $N \in [1000, 13000]$ subject to the constraints that $N\mu > y$ (to avoid data from the critical region $N\mu \simeq 0.15 \cdot y$) and $y < N/100$. The resulting values for the fit $z = d \cdot \mu^a y^b N^c$ (8) are

$$\begin{aligned} a &= 0.558(1), \quad b = 0.879(1), \\ c &= 0.091(1), \quad d = 1.79(2). \end{aligned} \quad (9)$$

We plot the turnover rate z against the form (4) using our best fit values (9) to create a data collapse. As it can be seen in the figure, collected data lay on the diagonal $z = \mu^a y^b N^c$. The changes in the coefficients of a and b for this large N data of (9) are statistically significant but not especially large. Using them to derive values for μ and y from a real data set is not likely to cause difficulties. The problem lies with the dependence on N , as the power c in N^c varies significantly and so leads

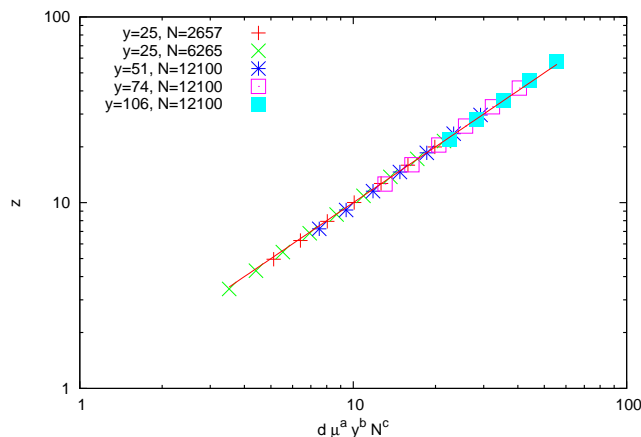


FIG. 11. Turnover rate z vs $d \cdot \mu^a y^b N^c$ for the Wright-Fisher model in the region $N\mu > y$, $y < N/100$. Values used were $a = 0.558$, $b = 0.879$, $c = 0.091$ and $d = 1.79$, the best fit values found (9). Error bars are smaller than symbols.

to large difference in numerical estimates from the fitted forms. These values for large N in (9) are again not compatible with the simple form (1) proposed in [10]. However the form $z = dy^a \mu^b N^c$ using the values derived from large N (7) now reproduce the data in table I for very large N within 6%. It should be noted, though, that in the $N \in [1000, 13000]$ region used for (9) results the dependence of the turnover rate z on the population size N is weaker than in the first fit performed. This might suggest a logarithmic dependence of the turnover rate z on the population size N as for $c \ll 1$: $N^c \simeq 1 + c \cdot \ln N$. Fits with this form work just as well as the one with the N^c dependence.

III. THE MORAN MODEL

In the Wright-Fisher model at each time step all individuals are assigned a new artefact simultaneously. While this can be a good update rule in some situations, it is certainly reasonable to think that cultural transmission and copying between individuals might occur in a more gradual way, with individuals changing taste individually one at a time.

This behaviour can be introduced in the model by changing the update rule so that at each time step only one randomly chosen individual is assigned a new artefact, by either copying it from an other individual in the population (chosen at random) or by inventing a new one with probability μ . This model is known as “Moran model” [2].

In this section the regularity of turnover in the top y chart of the most popular artefacts in the population for the Moran model is investigated as it has been done for the Wright-Fisher model in the previous section. The turnover z is defined as before and most of the analysis

Individual	t	t+1	t+2
1	A	A	D $\xrightarrow{\text{innovate}}$
2	A	A	A
3	B	A $\xrightarrow{\text{copy 1}}$	A
4	A	A	A
5	C	C	C
6	A	A	A

Top 3 chart:

Position	t	t+1	t+2
1 st	A	A	A
2 nd	B	C $\xrightarrow{z=1}$	C $\xrightarrow{z=1}$
3 rd	C		D

FIG. 12. A simple representation of the Wright-Fisher model. Shown are two successive time steps for a population of six individuals and the top chart at each time step.

is performed in the same way. Unless otherwise stated it is assumed that the investigation is conducted in the exact same way as for the Wright-Fisher model.

Steady state

Analytical results [2, 12] give that the typical time scale τ to reach a steady state is $\tau \sim N \min(\mu^{-1}, N)$, differing from that in the Wright-Fisher model by a factor N reflecting the different number of individuals it is possible to change each time step. This implies that simulations need to be run for a longer time with respect to the previous model. Again the steady state was confirmed numerically by studying $F_2(t)$ of (2) which evolves in time in the same way as in the Wright-Fisher model (3) but has a different eigenvalue controlling its evolution [2, 12] $\lambda_2 = 1 - 2(\mu/N) - 2(1 - \mu)/N^2$. For our parameter values we can safely choose $\tau = 4N\mu^{-1}$ as the starting point to compute \bar{z} . Simulations have been run for $T = 50 + N\mu^{-1}$ time steps (50 time steps have been added to ensure a minimum amount of time steps for big values of μ).

Data collection and analysis

The usual *ansatz* for the functional dependence of z on μ , y and N is made: $z = d \cdot \mu^a y^b N^c$ (4) and linear fits to $\ln(z)$ are used to estimate a , b , c and d . Approximately 350 data points have been collected in the following range $\mu \in [0.001, 0.357]$, $y \in [15, 256]$ and $N \in [100, 506]$. As figure 13 shows, there is a transition point in μ that separates two different behaviours of the turnover rate z . This critical point is roughly $\mu_c \simeq (N/y)^{-1.5}$ as can be seen in figure 14. In the region $\mu < \mu_c$ the observed be-

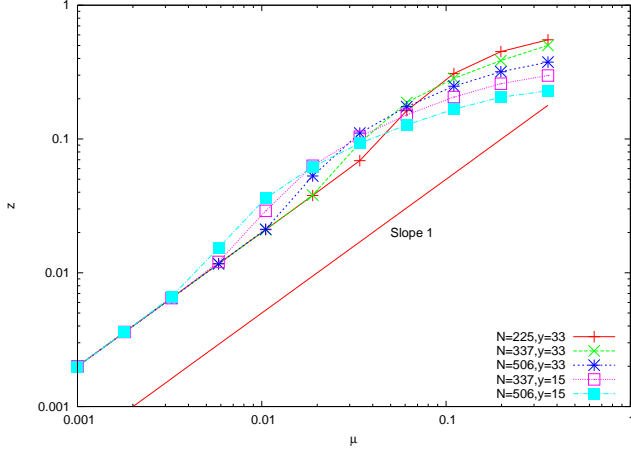


FIG. 13. Existence of a critical point. In the region $\mu < \mu_c$ we have $z \propto \mu$ (y and N are fixed for each of the plotted curves). Error bars are smaller than symbols.

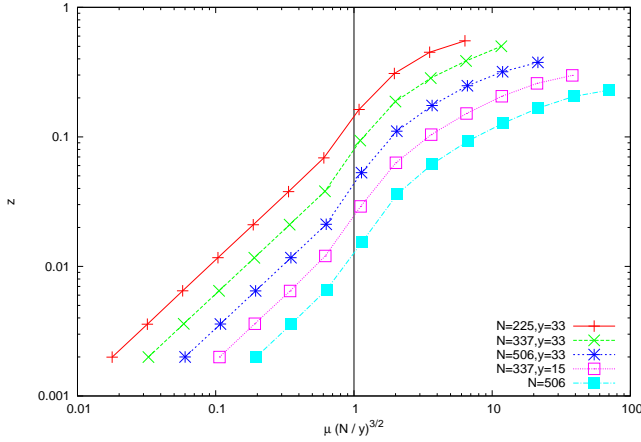


FIG. 14. Existence of a critical point $\mu_c \simeq (N/y)^{-3/2}$ (y and N are fixed for each of the plotted curves). Error bars are smaller than symbols.

behaviour of the turnover rate z is $z \propto \mu$, while for $\mu > \mu_c$ the turnover rate does not obey a simple power law as a function of the innovation rate μ .

$$\text{Region } \mu < \mu_c$$

Fitting the data points in this region with $z = d \cdot \mu^a y^b N^c$ (4) we find

$$\begin{aligned} a &= 0.9998(8), \quad b = -0.002(3), \\ c &= 0.001(3), \quad d = 2.02(3), \end{aligned} \quad (10)$$

so that $z = 2 \cdot \mu$ within one standard deviation. In this region z is independent on the top chart size y and the population size N . This is best seen in figure 13.

Keeping in mind that the average number of new artefacts introduced at each time step is equal to μ it is straightforward to understand that the mechanism that produces $z = 2 \cdot \mu$ is similar to the one described for the Wright-Fisher model: for $\mu \ll \mu_c$ a very small number of artefacts survives in the population in the steady state. It is then likely that the number of artefacts in the population at a certain time step t is lower than the top chart size y , as it is confirmed by observations (see figure 15). Suppose now to let the population evolve for N time steps, which is the unit of time in which all individuals in the population are assigned a new artefact. In this interval, on average, $N \cdot (1 - \mu)$ individuals will copy another artefact in the population, most likely one of the most popular ones, while $N\mu$ individuals will invent a new artefact. At this point we will have a few very popular artefacts in the population and the $N\mu$ new artefacts recently introduced, all in the top chart. If we now take other N time steps these very unpopular artefacts will likely be extinguished and replaced by other $N\mu$ invented artefacts. In the last N time steps $N\mu$ artefacts exit the top chart and $N\mu$ new artefact enter it. The average turnover in these N time steps is therefore $z = \frac{N\mu + N\mu}{N} = 2 \cdot \mu$

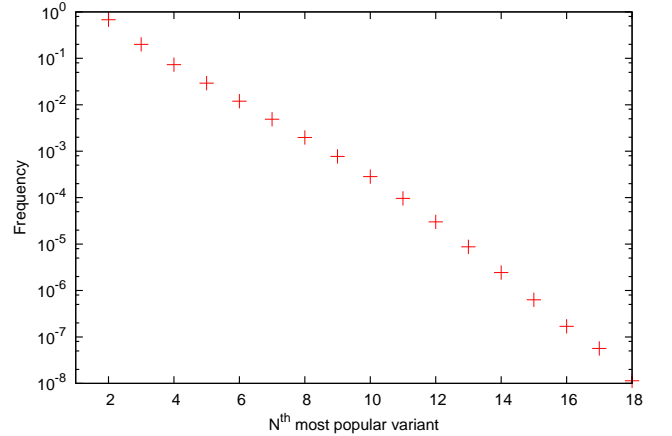


FIG. 15. Average frequency of artefacts (from the most popular one to the least popular). In the steady state there are no more than 18 artefacts in the population. With $N = 268$, $\mu = 0.00289$ and $y = 140$ we have $\mu \ll \mu_c \simeq (N/y)^{-3/2}$ and the maximum number of artefacts in the population (18) is smaller than the top chart size y (140).

$$\text{Region } \mu > \mu_c$$

We have seen in Fig.13, above the critical point the hypothesis of a power law dependence of the turnover rate z on the innovation rate μ is incorrect. Our results do suggest that the turnover rate z is a function of the ratio y/N , as can be seen in figure 16.

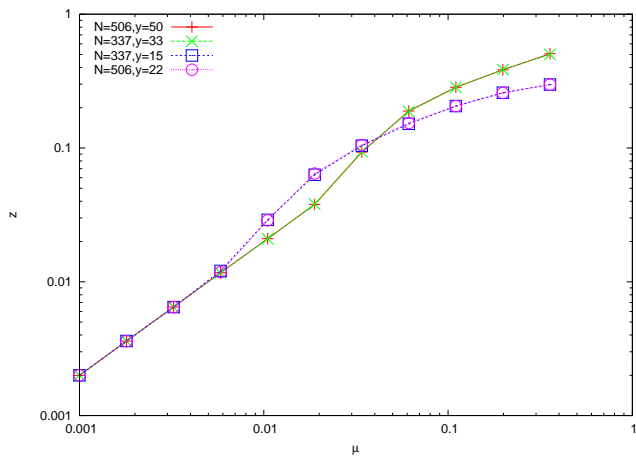


FIG. 16. z is a function of y/N : curves with the same value of the ratio y/N collapse. Error bars are smaller than symbols.

CONCLUSIONS

In this paper we have investigated the dependence of the turnover rate z on the parameters of the Wright-Fisher model and the Moran model.

We have found that in both models there is a critical point that separates two different regimes for the dependence of z on the innovation rate μ , the top chart size y and the population size N . This critical point satisfies the equation $N\mu \simeq 0.15 \cdot y$ for the Wright-Fisher model and $\mu \simeq (N/y)^{-3/2}$ for the Moran model. The two models behave similarly below the critical point, where the turnover rate is given by two times the average number of new artefacts entering the population in one time step. This is equal to $N\mu$ in the Wright-Fisher model and to μ in the Moran model.

Most data sets though will be above the critical region where we have seen that the two models behave differently, with the *ansatz* $z = d \cdot \mu^a y^b N^c$ being a good approximation for the functional dependence of the turnover rate on the model variables for the Wright-Fisher model only. The simple form $z = 2 \cdot \sqrt{\mu} \cdot y$ suggested by Bentley et al. [10] is excluded statistically by our results. We find that the powers of μ and y differ by 10% from the values in [10] and that there is significant dependence on the size of the system N . In particular we have also shown that our fit in the region where the top chart size is much smaller than the population size ($y \ll N$) (9) reproduces our simulation results for large populations within 6%, and we believe that this should be used when extracting information from real data sets. One outstanding issue is that our two fits (7) and (9) are not compatible in terms of the dependence of the turnover rate z on N . Further work is needed on this as this suggests the dependence of z on N may not be a simple power law.

For the Moran model for full top y charts, the usual

situation in real data, we have seen that the turnover doesn't follow a simple power law in μ , y and N . We have however noted that the turnover rate in the Moran model appears to be a function of the ratio y/N .

Since data on cultural transmission is sometimes available only in the form of popularity charts, the study of the turnover is of real practical use. However the models considered here and in [10] are only the simplest examples. It would be interesting to study the behaviour of the turnover in more complicated models such as examples which interpolate between the Wright-Fisher and Moran models [11, 13]. Another interesting avenue is to understand how the social network between individuals [13] alters the turnover.

ACKNOWLEDGEMENTS

We thank the High Performance Computing Centre at Imperial College London for use of their cluster in the large N studies.

APPENDIX

The information in this appendix is supplementary material which will not appear in the published version.

Wright-Fisher model, small N

For the Wright-Fisher model, we used $F_2(t)$ to check our numerical simulations reach a steady state on the time scale as can be seen in Fig. 17.

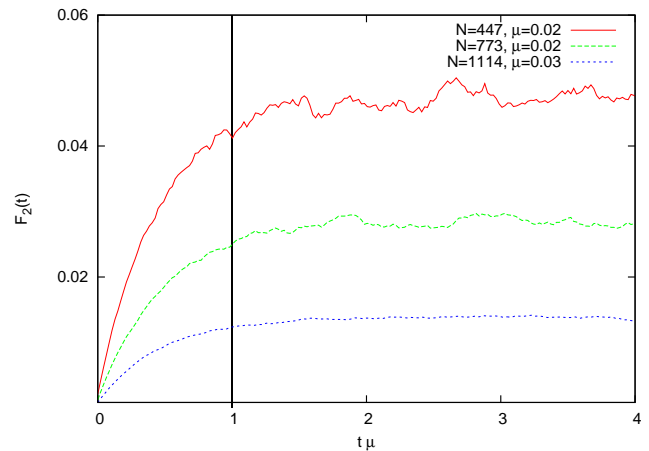


FIG. 17. Ensemble average of $\frac{\langle k(k-1) \rangle}{N(N-1)}$. The reach of a steady state can clearly be seen at $t \simeq 2\mu^{-1}$.

The observed approximate values of τ are listed in table II and compared with μ^{-1} in the two regions $N\mu < 1$ and $N\mu \geq 1$: it is found that the relation $\tau \sim \mu^{-1}$ holds

in the region $N\mu \geq 1$, while $\tau < \mu^{-1}$ for $N\mu < 1$. The observed behaviour of τ in the second region can be due to the fact that being less than one innovator per generation ($N\mu < 1$) on average, there is less variability in the artefacts and therefore a fastest approach to steady state through copying.

$\mu \backslash N$	50	100	500	1000	1500	μ^{-1}
0.02	50	50	50	50	50	50
0.01	80	100	100	100	100	100
0.002	250	300	500	500	500	500
0.001	300	400	800	1000	1000	1000
0.00067	300	300	1000	1500	1500	1500

TABLE II. Approximate values of τ . Above the main diagonal $N\mu \geq 1$, below $N\mu < 1$.

The data was fitted to (4) by applying the linear fit routine *fit* in *Gnuplot* to the logarithm of z and the variables μ , y and N

$$\ln z = a \cdot \ln \mu + b \cdot \ln y + c \cdot \ln N + \ln d. \quad (11)$$

on collected data points. For this purpose code and scripts have been set up to record $\ln \langle \bar{z} \rangle$ and $\sigma_{\ln \langle \bar{z} \rangle}$.

To perform the fit with equally spaced data μ , y and N are varied in the following way: $\mu = \mu_0 \cdot q^m$, $y = y_0 \cdot r^n$ and $N = N_0 \cdot s^p$ with fixed μ_0 , y_0 , N_0 , q , r , s and $m, n, p = 0, 1, 2, \dots$. We choose $q = 1.4$, $r = 1.3$, $s = 1.2$.

Moran model

We have from analytic results [2, 12] that the time scale for equilibrium is $\tau^{-1} \sim \ln(\lambda_2) \lesssim N \min((2\mu)^{-1}, N)$. This can be seen in our numerical simulations in figure 18.

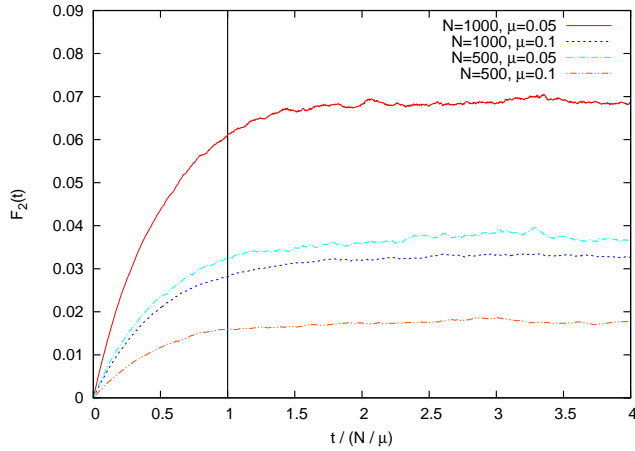


FIG. 18. Ensemble average of $\frac{\langle k(k-1) \rangle}{N(N-1)}$. The reach of a steady state can clearly be seen at $t \simeq 2N\mu^{-1}$.

To perform the fit with equally spaced data μ , y and N are varied in the following way: $\mu = \mu_0 \cdot q^m$, $y = y_0 \cdot r^n$ and $N = N_0 \cdot s^p$ with fixed μ_0 , y_0 , N_0 , $q = 1.5$, $r = 1.8$, $s = 1.5$ and $n = 0, 1, 2, \dots$. A further constraint is that the top chart size satisfies $y < N$.

-
- [1] G.A.Watterson, *The Annals of Mathematical Statistics*, **32**, 716–729 (1961).
 - [2] W.J.Ewens, *Mathematical Population Genetics: I. Theoretical Introduction*, (Springer-Verlag New York, 2004).
 - [3] C.Godrèche and J. M.Luck, “Nonequilibrium dynamics of urn models”, *J.Phys.Cond.Matter* **14** 1601 (2002).
 - [4] M.R.Evans and T.Hanne, “Nonequilibrium statistical mechanics of the zero-range process and related models”, *J.Phys.A.* **38** R195-R240 (2005).
 - [5] F.D.Neiman, “Stylistic variation in evolutionary perspective: Inferences from decorative diversity and inter-assemblage distance in Illinois Woodland Ceramic assemblages”, *American Antiquity* **60** 1 (1995).
 - [6] R.A.Bentley, H.D.G.Maschner, “Subtle criticality in popular album charts”, *Advances in Complex Systems* **2** 197–209 (1999)
 - [7] M.W. Hahn, R.A. Bentley, “Drift as a mechanism for cultural change: an example from baby names.”, *Proc.R.Soc.Lon. B* **270**, S120–S123 (2003).
 - [8] H.A. Herzog, R.A. Bentley, M.W. Hahn, “Random drift and large shifts in popularity of dog breeds.”, *Proc.R.Soc.Lon B (Supl.)* **271**, S353–S356 (2004).
 - [9] R.A. Bentley, M.W. Hahn, S.J. Shennan, “Random drift and culture change.”, *Proceedings of the Royal Society B* **271**, 1443–1450 (2004)
 - [10] R.A. Bentley, C.P. Lipo, H.A. Herzog, M.W. Hahn, “Regular rates of popular culture change reflect random copying.”, *Evolution and Human Behavior* **28**, 151–158 (2007)
 - [11] R.A.Bentley, P.Ormerod, and M.Batty, “An evolutionary model of long tailed distributions in the social sciences”., *Behav. Ecol. Sociobiol.* **65** 537-546 (2011)
 - [12] T.S. Evans, A.D.K. Plato, “Exact Solution for the Time Evolution of Network Rewiring Models.”, *Physical Review E* **75**, 056101 (2007)
 - [13] T.S.Evans, A.D.K.Plato and T.You., “Are Copying and Innovation Enough?”., in *Progress in Industrial Mathematics at ECMI 2008*, eds. Fitt, A. and Norbury, J. and Ockendon, H. and Wilson, E., **15**, Mathematics In Industry, 825–831, (2010) The European Consortium for Mathematics in Industry, Springer-Verlag
 - [14] T.S. Evans, *Exact solutions for network rewiring models*, *Eur. Phys. J. B* **56** (2007) 65.